グラフ構造に基づいたオンライン半教師あり学習に関する研究

著者	大堀 優
雑誌名	東北大学電通談話会記録
巻	86
号	1
ページ	234-235
発行年	2017-08
URL	http://hdl.handle.net/10097/00121448

修士学位論文要約（平成29年３月）

グラフ構造に基づいたオンライン半教師あり学習に関する研究

大堀 優　　指導教員：徳山 豪　　学位論文指導教員：全 眞嬉

A Study on Online Graph-based Semi-supervised Learning

Yu OHORI　　Supervisor: Takeshi TOKUYAMA　　Research Advisor: Jinhee CHUN

Semi-supervised learning using both labeled and unlabeled data is useful in situations where only a small subset of data is labeled. Online learning, which updates a learner with each observed data, is also important with the increase in demand for large scale data analysis. This study aims to provide an online semi-supervised learning method which adds a momentum term to online manifold regularization proposed by Goldberg et al. and experimentally show that the proposed method converges faster than the existing method.

1　序論

ラベル付きデータに加えてラベルなしデータを用いる半教師あり学習（SSL: Semi-Supervised Learning）は，少量のラベル付きデータしか得られない状況において，有用な学習方法である．また，データを観測する度に予測器を更新するオンライン学習は，昨今の大規模データ解析の需要増加に伴い，重要視されている．本研究では，2 つを組み合わせたオンライン半教師あり学習（OSSL: Online Semi-Supervised Learning）に取り組む．

SSL の一手法である Manifold Regularization (Batch MR) [1] を OSSL の枠組みに拡張した Online Manifold Regularization (Online MR) [2] では，予測器の更新にカーネルモデルに対する確率的勾配降下法（SGD: Stochastic Gradient Descent）[3] を用いている．SGD は汎用的手法であるものの，収束が遅いことが知られている [5]．本研究では，モーメンタム法を用いた Online MR を提案し，その収束速度について評価を行う．

2　問題設定

本研究では，半教師あり 2 クラス分類問題について議論する．今，入力列 $\{\boldsymbol{x}_t\}_{t=1}^T$ を考える．$\boldsymbol{x}_t \in \mathbb{R}^m$ は時刻 t に観測される入力である．このうち，l 個のデータにラベル $y_t \in \{\pm 1\}$ が付いているとする．このとき，汎化性能に優れた判別器 $h^\star$ を学習することがこの問題の目的である．判別器 h は，判別関数 f を用いて $h = \mathrm{sgn}\,(f)$ と表せる．

3　関連研究

ラベルなしデータを学習に利用するため，「データはある低次元多様体上に分布し，ラベルはその多様体上で滑らかに変化する」と仮定する．MR では，この仮定を満たす判別関数 $f^\star$ を求めるため，従来の正則化損失に新たな正則化項を加えた損失 $R_{\mathrm{reg}}(f)$ を最小化する．

$$R_{\mathrm{reg}}(f) = \frac{1}{l}\sum_{t=1}^{T}\delta(y_t)c(f(\boldsymbol{x}_t), y_t) + \frac{\lambda_1}{2}\|f\|_{\mathcal{H}}^2 + \frac{\lambda_2}{2T}\sum_{s,t=1}^{T}(f(\boldsymbol{x}_s) - f(\boldsymbol{x}_t))^2 W_{s,t} \tag{1}$$

ここで，$\delta(y_t)$ は入力 $\boldsymbol{x}_t$ にラベルが付いていれば 1，そうでなければ 0 を返す指示関数，c は任意の凸損失関数，$\|\cdot\|_{\mathcal{H}}$ は再生核ヒルベルト空間（RKHS: Reproducing Kernel Hilbert Space）$\mathcal{H}$ におけるノルム，$\lambda_1, \lambda_2 \geq 0$ は正則化の強さを調節する正則化パラメータ，$W_{s,t} \geq 0$ は入力 $\boldsymbol{x}_s, \boldsymbol{x}_t$ 間の類似度である．

続いて，時刻 t に観測される入力とラベルの組 $(\boldsymbol{x}_t, y_t)$ に対する損失 $R_{\mathrm{inst}}(\boldsymbol{x}_t, y_t, f)$ を $R_{\mathrm{reg}}(f) = \frac{1}{T}\sum_{t=1}^{T} R_{\mathrm{inst}}(\boldsymbol{x}_t, y_t, f)$ を満たすように定義する．

$$R_{\mathrm{inst}}(\boldsymbol{x}_t, y_t, f) = \frac{T}{l}\delta(y_t)c(f(\boldsymbol{x}_t), y_t) + \frac{\lambda_1}{2}\|f\|_{\mathcal{H}}^2 + \lambda_2\sum_{i=1}^{t-1}(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_t))^2 W_{i,t} \tag{2}$$

Online MR では，時刻 t にデータを観測する度に，カーネルモデルに対する SGD を用いて損失 $R_{\mathrm{inst}}(\boldsymbol{x}_t, y_t, f)$ が減少する方向へ判別関数 f を更新する．

$$f_{t+1} = f_t - \eta_t \left.\frac{\partial R_{\mathrm{inst}}(\boldsymbol{x}_t, y_t, f)}{\partial f}\right|_{f=f_t} \tag{3}$$

ここで，$\eta_t \in (0, \frac{1}{\lambda_1})$ は学習率といい，勾配降下の歩幅を表している．表現定理 [1] より，判別関数 f_t は特徴ベクトル $K(\boldsymbol{x}, \cdot) \in \mathcal{H}$ の線形和で表現できる．

$$f_t = \sum_{i=1}^{t-1}\alpha_i^{(t)} K(\boldsymbol{x}_i, \cdot) \tag{4}$$

ここで，K は任意の非負定値カーネル関数である．したがって，判別関数 f に関する更新式 (3) は，係数ベクトル $\boldsymbol{\alpha}$ に関する更新式 (5) に変形することができる．

$$\alpha_i^{(t+1)} = \begin{cases} (1-\eta_t\lambda_1)\alpha_i^{(t)} \\ \quad -2\eta_t\lambda_2(f_t(\boldsymbol{x}_i)-f_t(\boldsymbol{x}_t))W_{i,t} & i<t \\ 2\eta_t\lambda_2\sum_{i=1}^{t-1}(f_t(\boldsymbol{x}_i)-f_t(\boldsymbol{x}_t))W_{i,t} \\ \quad -\eta_t\frac{T}{l}\delta(y_t)c'(f_t(\boldsymbol{x}_t),y_t) & i=t \end{cases} \tag{5}$$

ここで，c' は凸損失関数 c の勾配である．

4　提案手法

本研究では，既存の Online MR の収束速度を向上させるため，モーメンタム法を用いた Online MR を提案する．はじめに，時刻 t における更新量 Δ_t を定義する．

$$\Delta_t = f_t - f_{t-1} = \sum_{i=1}^{t-1}\gamma_i^{(t)}K(\boldsymbol{x}_i,\cdot) \tag{6}$$

提案手法では，更新式 (3) の代わりに以下の更新式を用いる．

$$\begin{aligned} f_{t+1} &= f_t + \mu\Delta_t \\ &\quad -(1-\mu)\eta_t \left.\frac{\partial R_{\mathrm{inst}}(\boldsymbol{x}_t,y_t,f)}{\partial f}\right|_{f=f_t} \end{aligned} \tag{7}$$

これは，勾配に対して更新量の方向に慣性をつけたものと捉えることができる．ここで，$\mu\in[0,1)$ はモーメンタム係数といい，過去の更新量 Δ_t をどれだけ重視するかを表している．$\mu=0$ のとき，更新式 (7) は更新式 (3) と一致する．既存手法と同様に判別関数 f に関する更新式 (7) を変形すると，係数ベクトル $\boldsymbol{\alpha}$ に関する更新式 (8) が得られる．

$$\alpha_i^{(t+1)} = \begin{cases} \mu\gamma_i^{(t)} + (1-\eta_t\lambda_1)\alpha_i^{(t)} \\ \quad -2\eta_t\lambda_2(f_t(\boldsymbol{x}_i)-f_t(\boldsymbol{x}_t))W_{i,t} & i<t \\ 2\eta_t\lambda_2\displaystyle\sum_{i=1}^{t-1}(f_t(\boldsymbol{x}_i)-f_t(\boldsymbol{x}_t))W_{i,t} \\ \quad -\eta_t\dfrac{T}{l}\delta(y_t)c'(f_t(\boldsymbol{x}_t),y_t) & i=t \end{cases} \tag{8}$$

ここで，式を簡潔にするため，$(1-\mu)\eta_t$ を η_t とおき直している．最後に，係数ベクトル $\boldsymbol{\gamma}$ を求め，次回の更新に利用する．

$$\gamma_i^{(t+1)} = \begin{cases} \alpha_i^{(t+1)} - \alpha_i^{(t)} & i<t \\ \alpha_t^{(t+1)} & i=t \end{cases} \tag{9}$$

5　評価実験

提案手法の収束速度を評価するため，汎用的に使用される 10 のデータセットを対象に既存手法との比較実験を行った．データを予め標準化した後，以下の試行を 10 回繰り返し，その期待誤差の平均を算出した．公平性のため，ラベル付きデータの選択やシャッフルに用いる乱数のシードは，既存手法と提案手法で同じものを使用した．

(1) 訓練データのうち 2% をラベル付きデータとみなす
(2) 訓練データを無作為にシャッフルする
(3) 学習を行いながら一定の周期で期待誤差を測定する

既存手法，提案手法ともに類似度 $W_{s,t} = \exp(-\frac{\|\boldsymbol{x}_s-\boldsymbol{x}_t\|^2}{2\sigma^2})$，ヒンジ損失 $c(f(\boldsymbol{x}),y) = \max(1-yf(\boldsymbol{x}),0)$，ガウシアンカーネル $K(\boldsymbol{x}_i,\boldsymbol{x}_j) = \exp(-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{2\sigma_K^2}))$ を用いた．また，学習率を $\eta_t=\frac{0.1}{\sqrt{t}}$，提案手法のモーメンタム係数を $\mu=0.9$ とした．残りのハイパパラメータ $\gamma,\gamma_K,\lambda_1,\lambda_2$ はデータセットに応じて選択した．

実験の結果，10 のうち 9 のデータで既存手法を上回る性能を達成した．データセット Adult [4] における実験結果を図 1 に示す．

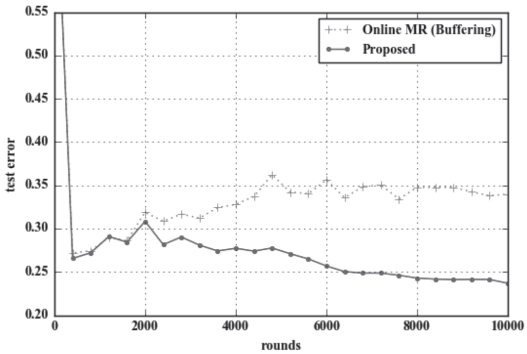

図 1　データセット Adult における実験結果

6　結論

本研究では，Online MR の収束速度を改善するため，モーメンタム法を用いた手法を提案した．また，既存手法との比較実験を行い，性能が向上することを確認した．今後の課題として，提案手法の理論解析が挙げられる．

参考文献

[1] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR 7*, 2006.
[2] A. B. Goldberg, M. Li, and X. Zhu. Online manifold regularization: A new learning setting and empirical study. In *ECML/PKDD*, 2008.
[3] J. Kivinen, A. J. Smola, and R. C. Williamson. Online learning with kernels. *IEEE Trans. Signal Processing 52(8)*, 2004.
[4] M. Lichman. UCI machine learning repository, 2013.
[5] 鈴木大慈. 確率的最適化. 機械学習プロフェッショナルシリーズ. 講談社, 2015.
[6] P. Vincent and Y. Bengio. Kernel matching pursuit. *Machine Learning 48(1-3)*, 2002.